DOCUMENT RESUME

ED 192 359 CS 205 874

AUTHOR Anderson, Beverly L.
TITLE Writing Assessment Alternative for the '80s.
PUB DATE Jul 80
NOTE 16p.: Paper presented at the Annual Meeting of the American Association of School Administrators (3rd, Chicago, IL, July 6-9, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Comparative Analysis; *Decision Making; *Educational Assessment; Educational Diagnosis; *Evaluation Methods; Formative Evaluation; Summative Evaluation; *Writing (Composition); Writing Research; Writing Skills
IDENTIFIERS *Holistic Scoring; *Writing Evaluation

ABSTRACT

This paper provides a brief overview of three direct measures of writing assessment--holistic, analytic, and primary trait scoring--and several variations and combinations of the three. It then relates these measures to eight types of educational decisions: (1) diagnosis of individual learner strengths and weaknesses, (2) student placement, (3) educational and vocational student guidance, (4) selection for admission to programs with enrollment limits, (5) certification of minimum competencies, (6) survey assessment, (7) formative program evaluation, and (8) summative program evaluation. (AEA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Beverly L. Anderson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED192359

Writing Assessment Alternatives for the '80s

Beverly L. Anderson*
Director, Clearinghouse for Applied Performance Testing (CAPT) and
Assessment and Measurement Program
Northwest Regional Educational Laboratory

Declining writing skills among our nation's students have received considerable attention recently by both educators and noneducators. States and local districts across the country have directed extensive financial and human resources toward determining the nature and extent of the problem, particularly through the assessment of writing skills. This paper highlights the main approaches being taken in assessing writing skill.

It is critical to keep in mind that there is not now, nor will there ever be, a single best way to assess writing skill. The method of choice is tied to the specific writing skills one desires to assess and the purpose for the assessment.

Following a brief overview of direct and indirect ways of assessing writing skills, the remainder of this paper focuses on the direct assessment of writing.

Two general methods of assessing writing proficiency are used--the direct and the indirect approach. The direct approach involves gathering samples of writing and evaluating them according to specified criteria. This method simulates real-life writing circumstances and requires trained judges to apply the criteria. The indirect approach involves the use of objective tests, such as paper-and-pencil multiple choice testing to

CS205274

*This article draws heavily from a book written by two CAPT staff members--Vicki Spandel and Richard Stiggins, entitled Direct Measures of Writing Skill: Issues and Applications (1980). It is available from the Office of Marketing, NWREL, 710 S.W. Second Avenue, Portland, OR 97204 for $4.25 (prepaid). The workshop presented at AASA was also based on materials prepared by Spandel and Stiggins.

measure language usage skills important to effective writing. The scoring of these assessments can generally be done by machine and thus is less costly.

Although the results of an indirect and direct assessment are generally highly related, indirect assessment should not be considered a substitute for direct assessment. Different skills are assessed: direct measures focus on writing composition skills while indirect measures focus on prerequisites of good writing. Thus, results provide different information.

The costs of conducting the two types of assessments differ. The cost of test scoring is high for direct assessment, while test development is the most costly feature of indirect assessment. Overall, indirect assessment is generally less expensive, partly because indirect measures are more readily available for purchase, precluding high development costs.

Developing a Direct Writing Assessment

In direct writing assessment, the examinee is given a set of instructions which defines the topic and conditions surrounding the writing assignment. For example, a writing exercise for a ninth grader may be*:

> Imagine that your local school board is considering adoption of a resolution to hold school year round. The opinions of everyone involved--including teachers, parents and students--will be considered. You have been asked, therefore, to write a letter to John Brown, chairman of the school board, stating your views.
>
> In your letter, state clearly whether you are for or against the 12-month school year. Give reasons for your opinion. Be specific. Use examples from your own experience to support your answer. Sign your letter "Pat Smith."

---

*This exercise is taken from the training materials now under development by CAPT.

The development of such exercises for direct assessment involves five steps:

- o assessment planning
- o exercise development
- o exercise review
- o exercise pretesting
- o final revision

The way and extent to which each step is carried out will vary depending on the assessment purpose and other conditions.*

Procedures for Scoring Writing Samples

While many objective tests can be machine scored, writing tests that rely on writing samples must be hand scored by persons trained to use designated criteria and performance standards. Several different methods have been devised for scoring writing samples, depending on the assessment purpose. Appropriateness depends upon what information is needed, how it will be used, and what resources are available. Some scoring methods are more complicated and costly than others.

Three scoring methods will be discussed here--holistic, analytic and primary trait scoring. The methods are similar in that all require careful training of scorers (usually language arts teachers) in a group setting for about half a day. Once scorers are trained, scoring is usually conducted in a carefully structured setting to allow reading of papers by two readers (and a resolver, if necessary) and exercising of

---

*A later section of this paper discusses assessment planning. See Spandel and Stiggins (1980) for details on the other steps. A consumer's guide to writing assessment now under development by CAPT will also address these issues.

quality control procedures. Under such conditions high reliability can be achieved.

Holistic Scoring

Holistic scoring involves reading a paper for an overall or "whole" impression. No specific trait, such as organization, syntax or originality, is individually addressed. A reader makes a judgment in much the same way that he or she decides whether a novel or an essay in Newsweek magazine is superior, mediocre, or slipshod.

Raters are asked to make no marks upon the paper. The objective is to score quickly, trusting first impressions.

Raters use "range finders" as a guide. These are actual student papers chosen as representative. There is a range finder for each of four score levels: i.e., 4, 3, 2 and 1. Virtually all trained readers agree on the score each paper would receive, so these papers serve as effective models to assist raters in assigning scores.

Range finder papers are intended to represent the approximate midpoint of the range. For example, a 4 range finder should not be considered the best possible paper from the sample. Rather, it should be considered a "middle 4"; that is, some 4's in the sample will be a little better, some not quite so good.

Range finders are intended as guides. Clearly, not every contingency can be anticipated. Readers are encouraged to (1) trust their own judgment, and (2) confer with a head reader to resolve any unusual problems.

One holistic score is determined to represent the overall quality of the writing. Although the specific strengths and weaknesses of a given

paper are not listed, a description of the characteristics usually found in papers receiving each score can accompany the score.

Analytical Scoring

Analytical scoring is a trait-by-trait analysis of a paper's merit. Individual traits considered important to any piece of writing in any context, are selected. For example, papers of students asked to write a letter expressing an opinion may be scored on ideas, organization and wording.*

The traits are scored one at a time. The scorer's impression of one trait should not influence the scoring of any other trait. Readers are presented with guidelines for scoring each trait. The guide presents an elaboration of each point on the rating scale for each trait. Scores are reported separately for each trait. Range finders such as those used in holistic scoring are sometimes used along with the scoring guides.

Primary Trait Scoring

Primary trait scoring is similar to analytical scoring in that it focuses on one or more specific characteristics of a given piece of writing. But they differ in that analytical scoring attempts to isolate general important characteristics, while primary trait analysis is rhetorically and situationally specific. The most important--or primary--trait(s) in a letter to the editor will not likely be the same as that (those) in a set of directions for assembling a toy.

---

*These are three of the five significant traits that researcher Paul Diederich isolated through his study. By soliciting feedback from experienced scorers trained in holistic procedures, Diederich (1974) isolated five characteristics which scorers seem to focus on in rating papers: ideas, mechanics, organization, wording and flavor.

The primary trait system is based on the premise that all writing is done in terms of an audience, and that successful writing will have the desired effect upon that audience, whether newspaper reader or employer. In scoring, papers are judged on the likelihood of their producing the desired response.

Because they are situation-specific, primary traits differ from item to item, depending on the nature of the assignment. Suppose a student were asked to give directions for taking the bus from home to school. The primary trait might then be sequential organization; any clear, unambiguous set of directions would necessarily be well organized. Table I summarizes the key features of the scoring procedures discussed above.

Variations and combinations of the methods can be appropriate in many circumstances. The scoring methods described have been refined and applied primarily in large-scale statewide or national assessment. When conducting writing assessment for use at the classroom or school level, consider these ideas:

1. Writing samples could be scored first using a holistic approach to provide an overall impression of the writing skills at a state or district level. Papers of those students receiving low scores could be scored later by teachers using an analytic or primary trait method to serve as a basis for planning remedial instruction.
2. Several teachers within a school could be trained in the scoring methods so that they could exchange student papers at certain times during the year. Consistency and communication in grading thus could be increased among teachers.
3. Students could be trained in the fundamentals of the scoring method to heighten their awareness of the components of high

0573D

TABLE I

COMPARISON OF ALTERNATIVES

	Holistic	Analytical	Primary Trait
GOAL	Overall Impression	Trait-by-trait Analysis	Situation Specific Analysis
PREPARATION	Allow 1/2 day to find range finders; 1/2 day to train readers	1 day to I.D. traits 1 day per trait to dev. criteria 1 day to refine scoring 1/2 day to train readers	Same as analytical
SCORING TIME	1 to 2 min. per paper depending on length	1 to 2 min. per trait per paper	1 to 2 min. per trait per paper
CONTEXT	Contexts where rank order is useful	Contexts where skill analysis is useful	Contexts where skill analysis is useful

quality writing. Peer scoring could be used as part of the instructional process.

4. For some purposes, it might be useful to train parents, employers, school board members and others in the scoring methods (holisticis likely to be the easiest one to use). For example, if a high school has a program in which students are placed in businesses as part of their school experience, employers' ratings of their writing could benefit both the student and the school.

Selecting the Appropriate Scoring Approach for the Assessment Purpose

The major criterion for selecting the scoring method should be the assessment purpose, since educational assessment has one function: to facilitate educational decision making.

Purposes for assessment can be divided into three broad categories--instructional management, entry-exit decisions and program planning and evaluation decisions.

Instructional Management

Tests can facilitate instructional management decisions. Data from these tests are used for (1) the diagnosis of individual learner strengths and weaknesses, (2) student placement, and (3) educational and vocational student guidance.

Diagnosis. Perhaps the most frequent reason for testing is to diagnose the strengths and weaknesses of individual students. In this context, the primary decision maker is the teacher, though the students themselves may also play a role in the decision. Teachers often use tests and other performance indicators to decide an individual student's current level of development and to select the next appropriate instructional unit for that student. Since tests that are useful in this decision are those

that reveal precisely what skills and knowledge the student has or has not mastered, an analytic or primary trait scoring is likely to be the most useful.

Placement. If diagnosis has as its goal the linking of students to individual instructional units within a course, placement can be thought of as linking the student to the next largest instructional element--the course or class. Placement tests usually cover a broader range of knowledge and skills than diagnostic tests and are only used once or twice a year, while diagnostic tests are used continually. A holistic score may serve this need.

Guidance. Diagnosis matches the student to an instructional unit, placement to a course, and guidance to an entire program of study. In this case, the primary decision makers are students and their parents, with assistance from guidance counselors. Each student must weigh the chances of success and satisfaction in various educational and vocational programs and then decide which to pursue. In terms of one's writing skill, a holistic score may be adequate for this purpose.

Entry-Exit Decisions

Tests are also administered in very structured circumstances to assist in making program entry or exit decisions about a student. For example, tests may be administered for the purpose of (1) selection for admission to programs with enrollment limits (e.g., college entrance or trade school), or (2) certification of minimum competencies (e.g., for high school graduation or occupational licensing).

Selection. The terms selection and placement are not always clearly differentiated. Placement, as described above, deals with the matching of all students tested to their appropriate level of instruction. This is an

instructional management decision. The term selection is used here to refer to those decisions where a limited number of spaces are available in an educational program and entrants are screened for admission based on which are most likely to benefit. The key decision makers are teachers and administrators of the selecting program. A test used for selection purposes (1) focuses on those skills and knowledge which are essential for future success, and (2) rank orders students in terms of relevant skills and knowledge to identify those most likely to succeed. College entrance and selection into a particular occupational training program are prime examples. A holistic score on a writing test may be used to rank order examinees to facilitate selection.

Certification. Tests often play an important role in certifying those who have attained a specific minimum acceptable level of educational development. Consider just three examples of certification involving different decision makers: First, a teacher might use a test to certify mastery of beginning verbal skills to decide whether to promote a given student. Second, a school district administrator acting in the name of the board of education might use a test to certify attainment of minimally acceptable skills for receipt of a high school diploma. And third, members of a certain technical profession might use a test to certify competence to practice that profession. In each case, examinees must pass the test to be certified as minimally competent. Therefore, the test must focus specifically on clearly stated minimal competencies. An analytic or primary trait score is likely to provide the most defensible score for this purpose.

Programmatic Decision

A third general use of tests is to facilitate educational programmatic decisions. Test data may be helpful in assessing the needs of a group to serve as the basis for developing a new program, allocating resources or evaluating existing programs. Such testing falls into three categories--(1) survey assessment, (2) formative program evaluation, and (3) summative program evaluation.

Survey assessment. One established use of testing in education is surveying student achievement and exploring trends over time to make programmatic planning decisions. This kind of testing is usually designed to raise questions rather than answer them. For example, why are writing scores gradually declining in the district (state and nation)? Or, why are scores of fourth graders consistently below national averages while those in other grades are above average? The test data are used to decide what aspects of the education system need to be more thoroughly investigated to determine possible reasons for unsatisfactory performance. For this purpose, achievement test scores--sometimes gathered from random samples of students--are generally gathered on an annual basis, then averaged across the entire school, district or state to show the general level of student developmment. Test scores are frequently compared from year to year to show trends. This information becomes a basis for setting educational policy and allocating resources. Typically educational administrators are the key decision makers, but they must justify their decisions to the taxpaying public. In essence, tests covering broad content and skill areas are used to conduct an overall assessment of an educational program. A holistic score is often sufficient to meet this need.

Formative evaluation. In formative evaluation, the goal is to determine which instructional units or features of a specific educational program (e.g., in remedial reading) are effective and which need revision. The decision makers are program teachers and administrators. Tests covering the specific interim and long-term outcomes of the program are administered during the program's operation to help shape the program during its formative stages. Most likely a primary trait or analytic score would be needed for this purpose.

Summative evaluation. Summative evaluation seeks to reveal a program's overall merit, thus suggesting whether that program should be continued or terminated. Tests designed to assess students' performance on final learning outcomes of a program are typically a key part of such an evaluation. Teachers, program, building or district administrators, and the public (including the board of education) may be involved in summative evaluation decisions. Tests may be given prior to as well as following instruction, with retention testing after a given interval. Depending on the program's goals, an analytic score would probably be best.

Thus, to gain an overall impression of a person's writing ability a holistic score may be sufficient. However, to determine the articular strengths and weaknesses of the person's writing ability (e.g., to serve as the basis for planning instructional programs), an analytic scoring method is warranted. A primary trait scoring system would be desired when one is concerned about a person's ability to write for a very specific purpose.

Summary

This paper provides a very brief overview of direct measures of writing assessment--its appropriate role, three score methods and

relationship to eight types of educational decisions. The reader is referred to Spandel and Stiggins (1980) for further discussion. The Clearinghouse for Applied Performance Testing, a National Institute of Education sponsored project, also is preparing a consumer's guide and training materials in writing assessment which will be available in the spring of 1981.

References

Spandel, V. and Stiggins, R. Direct Measures of Writing Skill: Issues and Applications. Portland, OR: Northwest Regional Educational Laboratory, 1980.

Diederich, P. Measuring Growth in English. Urbana, IL: National Council of Teachers of English, 1974.